REMARKS

Claims 1 and 3-23 were pending in the application. Claims 1 and 3-23 stand rejected. Claims 1, 6-7, 10-11, 17-18 and 20-21 were amended. Claims 24-29 were added. Claims 1 and 3-29 remain in the application.

Claims 1, 3-8, 10 and 23 stand rejected under 35 U.S.C. 103(a) as being unpatentable Qian et al., US Patent 6721454 (hereafter referred to as Qian 454),and further in view of Ratakonda, US Patent 5956026. The rejection stated:

'As in Claims 1 and 10, Qian et al. teaches a method and computer storage medium with instructions for obtaining unstructured video frames ("A video sequence 2 is input", Column 2, lines 64-65), generating segments from the shot boundaries based on the color dissimilarity between consecutive frames ("A color histogram technique may be used to detect the boundaries of the shots", Column 3, lines 42-43), extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61) for their visual dissimilarity and temporal relationship, and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ... events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8). While Qian et al. teaches extracting semantic events from unstructured video frames, they fail to show the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments as recited in the claims. In the same field of the invention, Ratakonda teaches a video event detection and segmentation merging method similar to that of Qian et al. In addition, Ratakonda further teaches the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments (Figures 1, 5 and corresponding text). It would have been obvious to one of ordinary skill in the art, having the teachings of Qian et al. and Ratakonda before him at the time the invention was made, to modify the segment generation and merging techniques taught by Qian et al. to include the processing of each pair of segments of Ratakonda, in order to obtain not only frames, but also inter-segment similarity processing. One would have been

-11-

motivated to make such a combination because layered hierarchical structure would have been obtained, as taught by Ratakonda.'

Claim 1 has been amended to state:

1. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure.

Amended Claim 1 is supported by the application as filed, notably the original claims and at page 12, line 21 to page 13, line 10.

Claim 1 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair" and "merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set. The cited references disclose no such metrics nor use of features including such metrics in a merging step. As noted in the rejection, Qian 454 teaches extracting semantic events from individual segments. Ratakonda discloses clustering of keyframes based upon application of one of several clustering algorithms to image histograms:

"In order to reconcile and satisfy diverse viewing requirements with the same video indexing system, a multi-resolution video browser, block 53, FIG. 2, is provided to allow a user to browse the hierarchical summary by selecting a specific level summary . This is a

-12-

browser instead of a mere indexing system. A viewer may start at a coarse level of detail and expand the detail with a mouse click at those parts of the keyframe sequence which are more interesting to the viewer. More than one level of detail is required so that the viewer may browse at a viewer-selected pace. The finest level keyframes still may be detected. At a coarser level, similar keyframes at the fine levels are clustered together and each cluster is represented by a representative keyframe.

"To solve this clustering problem, a modification of the well known Lindo-Buzo-Gray (LBG) algorithm (or Lloyd's algorithm or K-means algorithm)is proposed. Note that it is desirable to cluster similar images together. Assume that images are represented by their histograms and that similar images have similar histograms. Treating each histogram as a feature vector of its associated frame, find (N/r) representative histograms at the coarse level to replace the N histograms in the finest level, where N is the number of keyframes at the finest level."

(Ratakonda, col. 9, lines 30-53)

Ratakonda's clustering is unlike merging video segments with a merging criterion that applies a probabilistic analysis to the features of a feature set having an inter-segment temporal relationship feature that includes metrics of temporal separation between the segments and accumulated duration of the segments. A combination of the cited references would not address the untaught features.

The rejection argues as motivation for one or ordinary skill in the art to combine Qian 454 and Ratakonda:

"One would have been motivated to make such a combination because layered hierarchical structure would have been obtained, as taught by Ratakonda."

The references teach against use of the layered hierarchical structure of Ratakonda in Qian 454.

Ratakonda provides a method of hierarchical digital video summarization and browsing using a hierarchical summary based on keyframes. (Ratakonda, col. 2, lines 13-17) Keyframes are grouped at different levels:

"With the click of a button the user may access either the parent-children of the keyframe currently being viewed. Choosing the parent will result in the replacement of a group of keyframes at the current level by a single

keyframe which is their parent. Choosing the children will find all the child keyframes corresponding to the current keyframe. FIG. 5 illustrates this concept of parent and child keyframe. (Ratakonda, col. 13, lines 22-28)

"Tagging [i.e. clicking on] frames in the finest level 76 results in playback of the video". (Ratakonda, col. 5, lines 51-52; also see col. 5, lines 56-59)

Qian 454 also teaches against extracting inter-segment features by processing pairs of segments. In Qian 454, shots are compared in the form of summaries. Each of the individual shots, in Qian 454, are summarized with descriptors, such as "animal" and "tree", and the descriptors of different shots are compared, but not in pairs. (Qian 454, col. 10, line 61 to col. 12, line 9) Qian 454 teaches against comparisons between shots based upon "details" and teaches against presentation of image content to users. Qian 454 instead presents summaries to be <u>read</u> and interpreted:

"Each shot detected or forced at the first level 4 of the video content analysis technique is summarized 16 at the second level 12 of the technique. The shot summaries provide a means of encapsulating the details of the feature and motion analysis performed at the first 4 and second 12 levels of the technique so that an event inference module in the third level 18 of the technique may be developed independent of the details in the first two levels. <u>The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans. This facilitates</u> video indexing, retrieval, and browsing in video databases and the development of algorithms to perform these activities." (Qian 454, col. 10, line 63 to col. 11, line 6; <u>emphasis added</u>)

Ratakonda does not teach a layered hierarchical structure without the "details" of keyframes, thus, there is no motivation for a combination of the two references.

Claims 3-5, 9, and 23-25 are allowable as depending from Claim 1 and as follows.

The Office Action stated in relation to Claim 3:

"As in Claim 3, Qian et al. teaches obtaining unstructured video frames, generating segments from the shot boundaries based on the color dissimilarity between consecutive frames, extracting a set by processing pairs of segments for their visual dissimilarity and temporal

relationship by generating color histograms from the consecutive frames and from the histograms, generating a difference signal, thresholding of this signal based on a mean dissimilarity over several frames to produce a signal representative of the existence of a shot boundary (See Claim 23 rejection supra) and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure (See Claim 1 rejection supra) and the difference signal to be based on a mean dissimilarity over several frames centered on one frame. Qian et al. fails to teach basing the number of frames used to calculate the difference signal on a fraction of the frame rate of video capture as recited in the claims. Within the field of the invention, it would be obvious to one of ordinary skill in the art to base the number of frames on a fraction of the frame rate (See also Image Analysis and Mathematic Morphology, Vol. 1, Jean Serra). One would have been motivated to make such a combination because a shortened time frame for calculating the difference signal would have been obtained."

"Applicant has said that Claim 3 is allowable as depending from Claim 1, but has not addressed the obvious rejection of Claim 3, therefore the examiner asserts that it is an admission of prior art that within the field of the invention, it would be obvious to one of ordinary skill in the art to base the number of frames on a fraction of the frame rate (See above). The examiner assumes that the applicant acknowledges this rejection of obviousness. One would have been motivated to make such a combination because a shortened or lengthened (dependent upon the value of the fraction) time frame for calculating the difference signal would have been obtained."

Claim 3 has been broadened and now states:

> 3. The method as claimed in claim 23 wherein the difference signal is based on a mean dissimilarity determined over a plurality of frames centered on one of the consecutive frames.

Claim 3 requires a difference signal is based on a mean dissimilarity determined over a plurality of frames centered on one of the consecutive frames. Qian 454 and Ratakonda do not teach use of a plurality of frames centered on one of the consecutive frames for this purpose:

-15-

"A color histogram technique may be used to detect the boundaries of the shots. The difference between the histograms of two frames indicates a difference in the content of those frames. When the difference between the histograms for successive frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be used to detect the shot boundaries." (Qian 454, col. 3, lines 40-50; emphasis added)

"Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38 is performed using a threshold method, where differences between histograms of successive frames are compared." (Ratakonda, col. 4, lines 48-54)

The office action stated in relation to Claim 5:

"As in Claim 5, Qian et al. teaches computing a mean color histogram for each segment and a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments (Column 3, lines 42-50 and Figure 5)."

"In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra."

Claim 5 requires computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments. The rejection cites Qian 454 col. 3, lines 42-50. The cited portion of Qian 454 relates to determining the difference between histograms of frames to detect shot boundaries:

"A color histogram technique may be used to detect the boundaries of the shots. The <u>difference between the histograms of two frames</u> indicates a difference in the content of those frames. When the difference between the histograms for successive frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be

used to detect the shot boundaries." (Qian 454, col. 3, lines 40-50; emphasis added)

Ratakonda similarly states:

"Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38 is performed using a threshold method, where differences between histograms of successive frames are compared." (Ratakonda, col. 4, lines 48-)

The Office Action noted: "Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames".

The Office Action also cited Figure 5 of Qian 454 in relation to Claim 5. Figure 5 of Qian 454 relates to a "sample mean" that is unrelated to the subject matter of Claim 5. Qian 454 states:

"Referring to FIG. 5, in images with more than one moving object 60 and 62 an object's center position and size derived from the sample mean and standard deviation may be biased." (Qian 454, col. 6, lines 16-18)

Claim 6 has been amended to depend from Claim 7, which has been rewritten as an independent claim. Claims 6-8 are discussed below.

Claim 10 is supported and allowable on the grounds discussed below in relation to Claim 7.

Added Claim 24 states:

24. The method as claimed in claim 1 wherein said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments.

Claim 24 requires that the extracting of the inter-segment temporal relationship feature of each pair of segments includes determining a number of frames separating the respective pair of segments and determining an accumulated number of frames in the respective pair of segments. As noted in the rejection of

-17-

Claim 1, while Qian 454 teaches extracting semantic events from unstructured video frames, Qian 454 fails to show the generation of the inter-segment temporal relationship feature. Ratakonda also does not teach the determining steps of Claim 24. Ratakonda describes a hierarchical summary using keyframes:

"Referring now to FIG. 1, a hierarchical multilevel summary 20, which is generated by the hierarchical summarization method of the invention, may provide a detailed fine-level summary with sufficiently large numbers of frames, so that important content information is not lost, but at the same time provide less detailed summaries at coarser levels in order not to hinder the usage of a coarse or compact summary for fast browsing and identification of the video." (Ratakonda, col. 3, lines 30-37)

"A form for the hierarchical summary is depicted in FIG. 5, generally at 70. The hierarchical summary is divided into hierarchical keyframe levels." (Ratakonda, col. 5, lines 44-46)
"With the click of a button the user may access either the parent-children of the keyframe currently being viewed. Choosing the parent will result in the replacement of a group of keyframes at the current level by a single keyframe which is their parent. Choosing the children will find all the child keyframes corresponding to the current keyframe. FIG. 5 illustrates this concept of parent and child keyframe. (Ratakonda, col. 13, lines 22-28)
"Tagging [i.e. clicking on] frames in the finest level 76 results in playback of the video". (Ratakonda, col. 5, lines 51-52; also see col. 5, lines 56-59)

Added Claims 25-26 have language taken from Claims 7-8 and are also allowable on the same grounds as those claims. Claims 7-8 are discussed below.

Claim 7 was rewritten as an independent claim and now states:
7. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:
a) obtaining a plurality of frames of unstructured video;

-18-

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure;

wherein step d) comprises the steps of:

generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set, said parametric mixture models being statistical models; and

applying the merging criterion to the parametric mixture models.

The rejection stated in relation to Claim 7:

'As in Claim 7, Qian et al. teaches generating parametric mixture models (summaries created by shot summarization 16, Figure 1) to represent class-conditional densities of inter-segment features (based on temporal information and color analysis, See Claim 1 rejection supra) of the feature set and applying the merging criterion to the parametric mixture models (event inference 20/detected events 22, Figure 1).'

and

'In response to the arguments regarding claim 7, that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., "statistical models") are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See In re Van Geuns, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993).'

The language of Claim 7 makes explicit that parametric mixture models are statistical models. Amended Claim 7 requires:

"generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set, <u>wherein said parametric mixture models are statistical models</u>".

The combination of Qian 454 and Ratakonda does not disclose this feature. Qian 454 discloses summaries that are not statistical models. In Qian 454, summaries have shot descriptors, which are described as indicating as to a particular shot: "the existence of certain objects", "location and size information related to objects and the spatial relations between objects", and "motion information related to objects and the temporal relations between them". (Qian 454, col. 11, lines 9, 11-13, and 15-16) Ratakonda teaches a hierarchy of keyframes and groups of keyframes. (Ratakonda, Figures 1 and 5, col. 3, lines 30-45; col.5, lines 44-63; and col. 13, lines 22-31) The current rejection of Claim 7 relies upon an interpretation of "parametric mixture models" as not being statistical models:

'As in Claim 7, Qian 454 teaches generating parametric mixture models (summaries created by shot summarization 16, Figure 1)'.

'In response to the arguments regarding claim 7, that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., "statistical models") are not recited in the rejected claim(s).'

This rejection is overcome by amended Claim 7.

Claims 6 and 8 are allowable as depending from Claim 7 and as follows.

The rejection stated in relation to Claim 6:

'As in Claim 6, Qian et al. teaches processing pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration between pairs of segments ("each shot is summarized 16 ... events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).'

'In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra.'

Claim 6 was amended slightly grammatically. This change is supported in the same manner as Claim 1.

-20-

Claim 6 requires that the processing of pairs of segments includes processing for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments. The language relied upon in the rejection (Qian 454, col. 3, lines 6-8) relates to inferences based on <u>textual</u> shot summaries. Qian et al. states:

> "The shot summaries also abstract the lower level analysis results so that they can be <u>read</u> and interpreted more easily by humans." (Qian 454, col. 11, lines 1-3)

Qian 454 teaches summarization that encapsulates the details of the feature and motion analysis of each shot using descriptors. (Qian 454, col. 10, line 63 to col. 11, line 8) The domain specific event inference model uses the descriptors. (Qian 454, col. 11, lines 51-55) Events are inferred by matching the occurrence of objects and their spatial and temporal relationships detected in each of the shots. (Qian 454, col. 12, lines 6-7; generally see col. 11, line 58 to col. 12, line 9) Examples of shot descriptors are provided:

> 'In general, shot descriptors used in the shot summary include object, spatial, and temporal descriptors. The object descriptors indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc. The spatial descriptors represent location and size information related to objects and the spatial relations between objects in terms of spatial prepositions, such as "inside", "next to", "on top of", etc. Temporal descriptors represent motion information related to objects and the temporal relations between them. These may be expressed in temporal prepositions, such as, "while", "before", "after," etc.' (Qian 454, col. 11, lines 7-18)

Qian 454 does not teach descriptors of temporal separation between pairs of segments and/or for accumulated temporal duration between pairs of segments. In Qian 454, temporal descriptors represent motion information related to objects in a segment and the temporal relations between those objects <u>in that segment</u>. This is unlike Claim 6, which requires processing pairs of segments for a <u>temporal separation between</u> pairs of segments and for an <u>accumulated temporal duration of</u> pairs of segments.

Ratakonda does not disclose the features of Claim 6. Ratakonda uses histogram clustering to determine keyframes. (See Ratakonda, col. 9, lines 30-53, quoted above)

The Office Action stated in relation to Claim 8:

'As in Claim 8, it is notoriously well known that queues are used to implement hierarchical displays. The examiner takes official notice of this teaching. It would be obvious to one of ordinary skill in the art to combine the use of the organizing video segments into hierarchies with a queue implementation.'

'In response to the arguments regarding claim 8, Qian teaches the process of "inserting" merges frames together, constituting a pair of segments that define the event and updating the model of the merged segment. Ratakonda further illustrates step d as seen supra.'

The Office Action presents a different rejection of Claim 15, which is very similar to Claim 8:

"As in Claim 15, US Patent 6,721,454 and Ratakonda teach performing the merging in a hierarchical queue by initializing the queue by introducing each feature in the queue with a priority of the probability of merging each corresponding pair of segments, depleting the queue by merging the segments if the criterion is met, and updating the queue based on the updated model (See Claim 8 rejection supra)."

Clarification of the rejections of Claims 8 and 15 is requested, particularly as to the metes and bounds of the official notice taken and of the relied upon teachings of Qian 454 and Ratakonda. For the sake of advancing prosecution, it is assumed that the rejection of both Claims 8 and 15 is limited to what was stated in the rejection of Claim 8. The official notice taken is necessarily limited to its words. The rejection states that it is notoriously well known that queues are used to implement hierarchical displays. This statement addresses only one phrase of Claim 8: "performed in a hierarchical queue" and does not teach or suggest the steps of:

initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;

-22-

depleting the queue by merging the segments if the

merging criterion is met; and

updating the model of the merged segment and then

updating the queue based upon the updated model.

The rejection also does not teach or suggest performance of step d (that is, merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure) using the above steps relative to the hierarchical queue. What the official notice taken does not address cannot be considered to be taught or suggested. MPEP 2144.03 states:

"If such notice is taken, the basis for such reasoning must be set forth

explicitly. The examiner must provide specific factual findings

predicated on sound technical and scientific reasoning to support his

or her conclusion of common knowledge. See Soli, 317 F.2d at 946,

37 USPQ at 801; Chevenard, 139 F.2d at 713, 60 USPQ at 241."

The rejection will not stand and must be withdrawn.

Claims 9 and 11-22 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., US Patent 6,721,454 (hereafter (Qian 454) and Ratakonda, US Patent 5,956,026 and further in view of Qian et al., US Patent 6,616,529 (hereafter Qian 529). The rejection stated:

'As in Claims 9, 11, 17-18 and 20, US Patent 6,721,454 and

Ratakonda teach a method and computer storage medium with instructions

for obtaining unstructured video frames, generating segments from the

shot boundaries based on the color dissimilarity between consecutive

frames, extracting a set by processing pairs of segments for their color

dissimilarity and temporal relationship of each pair of segments, merging

adjacent video segments by applying a probabilistic analysis to the

extracted set to represent the video structure, and generating a parametric

mixture model of the inter-segment features (See Claim 1 rejection supra).

While US Patent 6721454 and Ratakonda teach the segmentation due to

color dissimilarity, extraction due to visual dissimilarity and temporal

relationships, merging with probabilistic analysis and generation of a

parametric mixture model, they fail to show the probabilistic analysis to be

a Bayesian analysis applied to the parametric mixture model, and

-23-

representing the merging sequence in a hierarchical tree structure as recited in the claims. US Patent 6,616,529 teaches a video segmentation method similar to that of US Patent 6,721,454 and Ratakonda. In addition, US Patent 6,616,529 further teaches the probabilistic analysis to be a Bayesian analysis applied to the parametric mixture model (Figure 3 and corresponding text in Columns 4-5), and representing the merging sequence in a hierarchical tree structure (Figures 2a-2g and corresponding text). It would have been obvious to one of ordinary skill in the art, having the teachings of US Patent 6,721,454 and Ratakonda and US Patent 6,616,529 before him at the time the invention was made, to modify the segmentation with color dissimilarity and temporal relationships with a parametric mixture model taught by US Patent 6,721,454 and Ratakonda to include the construction of hierarchy according to probabilistic merging with Bayesian analysis of US Patent 6,616,529, in order to obtain a hierarchical representation of the frames grouped by color dissimilarity and temporal relationships according to Bayesian probability methods of analysis. One would have been motivated to make such a combination because a visual representation of the segmented video would have been obtained, as taught by US Patent 6,616,529 (Column 2, lines 24-55).

Claim 9 is allowable as depending from Claim 1.

Claim 11 has been amended to state:

11. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model; and

e) merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure.

Claim 11 requires "said parametric mixture model being a statistical model" and is supported and allowable as discussed above in relation to Claim 7.

Claims 12-16 are allowable as depending from Claim 11 and as follows.

Claims 12-13 are also allowable on the same basis as Claims 5-6, respectively.

Claim 14 requires that the parametric mixture model that is a statistical model and is generated in step d) represents class-conditional densities of the inter-segment features comprising the feature set. This is not disclosed by the cited references.

Claim 15 is allowable on the same basis as Claim 8.

Amended Claim 17 states:

17. A computer storage medium having instructions stored therein for causing a computer to perform acts for structuring video by probabilistic merging of video segments, the acts including:

obtaining a plurality of frames of unstructured video;

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

generating a parametric mixture model of the inter-segment features comprising the feature set, said parametric mixture model being a statistical model; and

merging video segments with a merging criterion that applies a probabilistic Bayesian analysis to the parametric mixture model, thereby generating a merging sequence representing the video structure.

Claim 17 requires "said parametric mixture model being a statistical model" and is supported and allowable as discussed above in relation to Claim 7.

Amended Claim 18 states:

18. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived from the feature set, said parametric mixture models being statistical models, thereby generating a merging sequence; and

e) representing the merging sequence in a hierarchical tree structure.

Claim 18 requires "said parametric mixture models being statistical models" and is supported and allowable as discussed above in relation to Claim 7.

Claim 19 is allowable as depending from Claim 18.

Amended Claim 20 states:

20. A computer storage medium having instructions stored therein for causing a computer to perform probabilistic merging of video segments, said instructions performing the acts of:

a) obtaining a plurality of frames of unstructured video;

b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive video frames;

c) extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments;

d) merging adjacent video segments with a merging criterion that applies a probabilistic Bayesian analysis to parametric mixture models derived from the feature set, said parametric mixture models being a statistical models, thereby generating a merging sequence; and

e) representing the merging sequence in a hierarchical tree structure.

Claim 11 requires "said parametric mixture models being statistical models" and is supported and allowable as discussed above in relation to Claim 7.

Amended Claim 21 states:

21. A method for structuring video by probabilistic merging of video segments, said method comprising:

generating video segments from an unstructured plurality of video frames by detecting shot boundaries based on color dissimilarity between consecutive frames;

extracting a feature set by processing pairs of segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair;

merging the video segments with a merging criterion that applies a probabilistic analysis to the feature set, thereby generating a merging sequence representing the video structure, the merging being independent of any empirical parameter determination; and

generating a hierarchy with the merged video segments, the hierarchy having a merging sequence represented by a binary partition tree.

A typing error was corrected. Claim 21 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair" and is supported and allowable on the same basis as Claim 1.

-27-

Claim 22 is allowable as depending from Claim 21.

Added Claim 27 states:

27. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

generating video segments from a plurality of frames of unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

computing an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to said features, thereby generating a merging sequence representing the video structure.

Claim 27 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair" and is supported and allowable on grounds discussed above in relation to Claim 1.

Added Claim 28 states:

28. The method of claim 27 wherein said computing of said inter-segment temporal relationship feature of each said pair of segments further comprises determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments.

Claim 28 is supported and allowable on grounds discussed above in relation to Claim 24.

Claim 29 states:

29. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:

obtaining a plurality of frames of unstructured video;

-28-

generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

extracting an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments, said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments; and

merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure.

Claim 29 requires "said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments" and is supported and allowable on the same grounds as Claims 27.

It is believed that these changes now make the claims clear and definite and, if there are any problems with these changes, Applicants' attorney would appreciate a telephone call.

In view of the foregoing, it is believed none of the references, taken singly or in combination, disclose the claimed invention. Accordingly, this application is believed to be in condition for allowance, the notice of which is respectfully requested.

Respectfully submitted,

Attorney for Applicant(s)
Registration No. 30,700

Robert Luke Walker/amb
Rochester, NY 14650
Telephone: (585) 588-2739
Facsimile: (585) 477-1148

If the Examiner is unable to reach the Applicant(s) Attorney at the telephone number provided, the Examiner is requested to communicate with Eastman Kodak Company Patent Operations at (585) 477-4656.